

TESIS

MODIFIKASI ALGORITMA *J-BIT ENCODING* UNTUK
MENINGKATKAN RASIO KOMPRESI



Johanes K. M. Lobang
No. Mhs. : 135302016/PS/MTF

PROGRAM STUDI MAGISTER TEKNIK INFORMATIKA
PROGRAM PASCASARJANA
UNIVERSITAS ATMA JAYA YOGYAKARTA

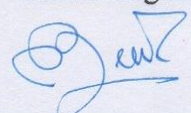

2016



UNIVERSITAS ATMA JAYA YOGYAKARTA
PROGRAM PASCASARJANA
PROGRAM STUDI MAGISTER TEKNIK INFORMATIKA

PERSETUJUAN TESIS

Nama : Johanes K. M. Lobang
Nomor Mahasiswa : 135302016/PS/MTF
Konsentrasi : *Soft Computing*
Judul tesis : Modifikasi Algoritma *J-Bit Encoding* Untuk
Meningkatkan Rasio Kompresi

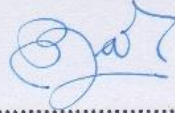
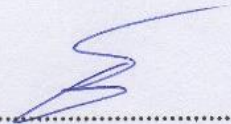
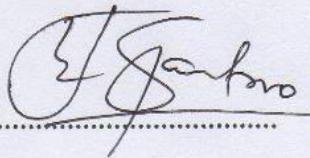
Nama Pembimbing	Tanggal	Tanda tangan
Dr. Pranowo, S.T., M.T.	19-12-2016	
Prof. Ir. Suyoto, M.Sc., Ph.D	19-12-2016	



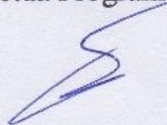
UNIVERSITAS ATMA JAYA YOGYAKARTA
PROGRAM PASCASARJANA
PROGRAM STUDI MAGISTER TEKNIK INFORMATIKA

PENGESAHAN TESIS

Nama : Johanes K. M. Lobang
Nomor Mahasiswa : 135302016/PS/MTF
Konsentrasi : *Soft Computing*
Judul tesis : Modifikasi Algoritma *J-Bit Encoding* Untuk
Meningkatkan Rasio Kompresi

Nama Penguji	Tanggal	Tanda tangan
Dr. Pranowo, S.T., M.T. (Ketua)	19-12-2016	
Prof. Ir. Suyoto, M.Sc., Ph.D (Sekretaris)	19-12-2016	
Dr. Ir. Alb. Joko Santoso, M.T. (Anggota)	19-12-2016	

Ketua Program Studi


Prof. Ir. Suyoto, M.Sc., Ph.D

PERNYATAAN

Dengan ini penulis menyatakan bahwa dalam tesis ini tidak pernah terdapat karya yang pernah diajukan untuk memperoleh gelar kesarjanaan di suatu Perguruan Tinggi, dan sepanjang pengetahuan penulis juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan di dalam daftar pustaka.

Demikian pernyataan ini dibuat untuk digunakan sebagaimana mestinya.

Yogyakarta, November 2016

Yang membuat pernyataan

Johanes K. M. Lobang

KATA PENGANTAR

Penulis mengucapkan puji syukur kepada Tuhan Yang Maha Esa karena atas perkenaan-Nya sehingga tesis ini dapat diselesaikan. Tujuan dari pembuatan tesis ini adalah sebagai salah satu syarat untuk mencapai derajat Strata Dua (S-2) pada Program Studi Magister Teknik Informatika Universitas Atma Jaya Yogyakarta.

Penulis menyadari bahwa terselesaikannya tesis ini tidak terlepas dari bantuan berbagai pihak baik secara langsung maupun tidak langsung. Untuk itu, penulis mengucapkan terima kasih kepada:

1. Bapak Dr. Pranowo, S.T., M.T. selaku dosen pembimbing I.
2. Bapak Prof. Ir. Suyoto, M.Sc., Ph.D selaku dosen pembimbing II dan Ketua Program Studi Magister Teknik Informatika.
3. Bapak Dr. Ir. Alb. Joko Santoso, M.T. selaku dosen penguji.
4. Universitas Nusa Cendana Kupang yang telah memberikan bantuan beasiswa.
5. Teman-teman angkatan 2013 dan 2014 Universitas Atma Jaya Yogyakarta.
6. Orang tua, isteri, anak-anak dan keluargaku.
7. Warga Kutu Dukuh dan Popongan - Sinduadi – Mlati – Sleman.
8. Umat lingkungan St. Demetrius Paroki St. Alfonsus Nandan.

Akhir kata penulis berharap semoga tesis ini dapat bermanfaat bagi pembaca walaupun masih banyak kekurangannya.

Yogyakarta, November 2016

Penulis,

Johanes K. M. Lobang

INTISARI

J-bit encoding merupakan algoritma kompresi *lossless* yang memanipulasi setiap bit data dalam file untuk meminimalkan ukuran, dengan cara membagi data menjadi dua output kemudian dikombinasikan kembali menjadi satu output. Penelitian ini mengusulkan modifikasi algoritma *J-bit encoding* dengan cara mengeliminasi simbol nol dan satu dari output pertama, sehingga output pertama akan berisi data asli selain nol dan satu (dalam ukuran *byte*) dan output kedua akan berisi nilai dua bit yang menjelaskan posisi *byte* nol, *byte* satu, dan *byte* selain nol dan satu. Perbandingan kedua algoritma ini dilakukan dengan menguji empat skema kombinasi algoritma yaitu (i) transformasi Burrows-Wheeler, *Move to Front*, *J-bit encoding* dan pengkodean aritmatika, (ii) transformasi Burrows-Wheeler, *Move to Front*, algoritma hasil modifikasi dan pengkodean aritmatika, (iii) transformasi Burrows-Wheeler, *Move One From Front*, *J-bit encoding* dan pengkodean aritmatika, (iv) transformasi Burrows-Wheeler, *Move One From Front*, algoritma hasil modifikasi dan pengkodean aritmatika. Dengan menggunakan *dataset* calgary corpus dan canterbury corpus, hasil pengujian menunjukkan bahwa rata-rata rasio kompresi terbaik diperoleh dengan menggunakan skema kedua. Selain efektif, algoritma hasil modifikasi juga lebih efisien dibandingkan dengan algoritma *J-bit encoding*.

Kata-kata kunci: Kompresi Data, Burrows-Wheeler *Compression Algorithm*, *J-Bit Encoding*

ABSTRACT

J-bit encoding is a lossless data compression algorithm that manipulates each bit of the file data in order to minimize the size by dividing the data into two outputs and combining the data into two outputs. This research proposes a modification of the J-bit encoding algorithm by eliminating zero and one symbols of the first output, so that the first output will contain the original data without zero and one symbols and the second output will contain the value of two bits that describe the position of zero, one, and byte besides zero and one. Comparison of the two algorithms is done by testing four scheme combination algorithm, namely (i) Burrows-Wheeler transformation, Move to Front, J-bit encoding and arithmetic coding, (ii) Burrows-Wheeler transformation, Move to Front, modification of the J-bit encoding and arithmetic coding, (iii) Burrows-Wheeler transformation, Move One From Front, J-bit encoding and arithmetic coding, (iv) Burrows-Wheeler transformation, Move One From Front, modification of the J-bit encoding and arithmetic coding. By using the calgary corpus and canterbury corpus datasets, the test results showed that the best compression ratio is obtained by using a second scheme on average. In addition to effective, modification of the J-bit encoding is also more efficient than the J-bit encoding algorithm.

Keywords: Data Compression, Burrows-Wheeler Compression Algorithm, J-Bit

Encoding

DAFTAR ISI

HALAMAN JUDUL.....	i
HALAMAN PERSETUJUAN TESIS	ii
HALAMAN PENGESAHAN TESIS	iii
PERNYATAAN.....	iv
KATA PENGANTAR	v
INTISARI.....	vi
ABSTRACT.....	vii
DAFTAR ISI.....	viii
DAFTAR TABEL.....	x
DAFTAR GAMBAR	xi
DAFTAR LAMPIRAN.....	xii
ARTI LAMBANG DAN SINGKATAN	xiii
BAB I PENDAHULUAN	1
A. Latar Belakang.....	1
B. Perumusan Masalah	3
C. Batasan Masalah	3
D. Manfaat Penelitian	4
E. Tujuan Penelitian	4
F. Sistematika Penulisan	4
BAB II TINJAUAN PUSTAKA.....	6
A. Sejarah Kompresi Data	6
B. Penelitian Tentang BWCA dan JBE.....	8
BAB III LANDASAN TEORI.....	12
A. Kompresi Data	12
B. Burrows-Wheeler Compression Algorithm (BWCA)	13
1. Transformasi Burrows-Wheeler (BWT)	14
2. <i>Global Structure Transform</i> (GST).....	16
3. <i>Run-Length Encoding</i> (RLE)	18
4. <i>Entropy Encoding</i> (EC).....	19
C. <i>J-Bit Encoding</i> (JBE).....	20
1. Proses Kompresi	20

2. Proses Dekompresi:	21
BAB IV METODOLOGI PENELITIAN	22
A. Alat dan Bahan	22
B. Metode yang digunakan.....	24
C. Langkah-langkah Penelitian	24
BAB V ANALISIS DAN MODIFIKASI ALGORITMA JBE	26
A. Analisis Data Uji.....	26
B. Modifikasi Algoritma JBE.....	27
C. Analisis Perbandingan Rasio Kompresi	28
D. Algoritma Hasil Modifikasi	30
1. Proses kompresi	30
2. Proses dekompresi	31
BAB VI HASIL PENELITIAN DAN PEMBAHASAN	33
A. Implementasi Sistem.....	33
B. Analisis Hasil.....	36
BAB VII KESIMPULAN DAN SARAN	42
A. Kesimpulan	42
B. Saran	43
DAFTAR PUSTAKA	44
LAMPIRAN	48

DAFTAR TABEL

Tabel 2.1 Kinerja BWT Dibandingkan Algoritma Lainnya	9
Tabel 3.1 Proses BWT pada String “RAKSASA”	15
Tabel 3.2 Proses Encoding MTF.....	17
Tabel 3.3 Proses Encoding M1FF.....	17
Tabel 3.4 Proses Decoding MTF	18
Tabel 3.5 Proses Decoding M1FF.....	18
Tabel 4.1 <i>Dataset</i> Calgary Corpus	22
Tabel 4.1 <i>Dataset</i> Calgary Corpus lanjutan	23
Tabel 4.2 <i>Dataset</i> Canterbury Corpus.....	23
Tabel 4.2 <i>Dataset</i> Canterbury Corpus lanjutan.....	24
Tabel 5.1 Frekuensi Simbol dari <i>Dataset</i> Calgary Corpus	26
Tabel 5.2 Frekuensi Simbol dari <i>Dataset</i> Canterbury Corpus	27
Tabel 5.3 Perbandingan Rasio Kompresi.....	29
Tabel 5.3 Perbandingan Rasio Kompresi Lanjutan	30
Tabel 6.1 File-File Hasil Implementasi Sistem.....	33
Tabel 6.2 Perbandingan Rasio Kompresi pada <i>Dataset</i> Calgary Corpus.....	37
Tabel 6.3 Perbandingan Rasio Kompresi pada <i>Dataset</i> Canterbury Corpus	38

DAFTAR GAMBAR

Gambar 3.1 Skema Umum BWCA.....	13
Gambar 4.1. Bagan Alir Langkah-Langkah Penelitian.....	25
Gambar 5.1 Gambaran Proses Kompresi Algoritma Hasil Modifikasi.....	32
Gambar 6.1 Menu Utama Aplikasi Kompresi Data.....	34
Gambar 6.2 Antar Muka Proses Kompresi Data.....	35
Gambar 6.3 Antar Muka Proses Dekompresi Data.....	35
Gambar 6.4 Perbandingan Waktu Proses Kompresi (Dalam Detik).....	40
Gambar 6.5 Perbandingan Waktu Proses Dekompresi (Dalam Detik).....	41

DAFTAR LAMPIRAN

A. <i>Source Code</i> Program	48
1. bwt.h.....	48
2. bwt.c	49
3. bitfile.h	52
4. bitfile.c	54
5. arcode.h.....	69
6. arcode.c	70
7. jbit1.h	83
8. jbit1.c.....	84
9. mainprogram.c	96

ARTI LAMBANG DAN SINGKATAN



AC	= <i>Arithmetic Coding</i> / pengkodean aritmatika
AES	= <i>Advanced Encryption Standard</i>
ASCII	= <i>American Standard Code for Information Interchange</i>
BWCA	= <i>Burrows-Wheeler Compression Algorithm</i>
BWT	= <i>Burrows-Wheeler Transform</i>
CM	= <i>Context Mixing</i>
DC	= <i>Distance Coding</i>
EC	= <i>Entropy Coding</i>
GST	= <i>Global Structure Transform</i>
IF	= <i>Inversion Frequencies</i>
IFC	= <i>Incremental Frequency Count</i>
LZ77	= Lempel-Ziv 77
LZ78	= Lempel-Ziv 78
LZW	= Lempel-Ziv Welch
JBE	= <i>J-Bit Encoding</i>
M1FF	= <i>Move One From Front</i>
MTF	= <i>Move To Front</i>
PPM	= <i>Prediction by Partial Matching</i>
RLE	= <i>Run-Length Encoding</i>